Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 10(1) – Year: 2026 – Pages: 56-78

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.1600372

Received: 12.12.2024 ***Accepted***: 04.03.2025 *Research Article*

Proteasome Inhibition by Biflavonoids from the Genus Araucaria: Insights from 3D-QSAR Modeling, Molecular Docking and Molecular Dynamics Simulation

Fahruddin Hisanurrijal, Purwantiningsih Sugita[1], Luthfan Irfana

Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia.

Abstract: Biflavonoid compounds have demonstrated significant potential as anticancer agents, particularly as non-covalent proteasome inhibitors. However, the inhibitory mechanisms of these compounds remain underexplored. The 20S proteasome, a key target in cancer therapy, plays a crucial role in protein degradation and cell cycle regulation, making its inhibition a promising strategy for cancer treatment. This study employs an integrated computational approach, combining Three-Dimensional Quantitative Structure-Activity Relationships (3D-QSAR) modelling, molecular docking, molecular dynamics (MD) simulations, and Molecular Mechanics-Generalized Born and Surface Area Solvation (MM/GBSA) binding energy calculations, to evaluate the proteasome inhibitory potential of biflavonoids from the genus *Araucaria*. A 3D-QSAR model was developed using 62 flavonoid derivatives, with the Partial Least Squares (PLS) model highlighting electrostatic interactions and hydrogen bond donors as key determinants of proteasome inhibition. Concurrently, the Support Vector Machine (SVM) model exhibited superior predictive accuracy (with an R^2 of 0.98 and a predicted R^2 of 0.75) and was employed to screen 22 biflavonoid compounds, identifying five candidates with the highest predicted IC_{50} values: 7-*O*-methylagathisflavone **(1)**, 7-*O*-methylcupressuflavone (**15**), ochnaflavone (**22**), 7"-*O*-methylamentoflavone (**11**), and 7"-*O*-methylagathisflavone (**2**). Molecular docking analysis confirmed strong binding affinities of all five compounds within the β5 active site of the 20S proteasome, with **22** exhibiting the highest docking score. However, MD simulations (100 ns) provided a more comprehensive assessment of binding stability, revealing that **1** showed the most stable behaviour, characterized by low RMSD fluctuations, minimal RMSF values, and a stable radius of gyration (Rg). Conversely, **15** and **22** demonstrated substantial conformational fluctuations, indicating diminished long-term stability. MM/GBSA binding energy calculations further substantiated the ranking observed in 3D-QSAR predictions, underscoring the preeminent potential of compound **1** as a promising inhibitor, as it demonstrated the best IC_{50} prediction, the strongest binding interactions, and the highest dynamic stability. The integration of 3D-QSAR modelling, docking, and MD simulations provides a comprehensive evaluation of biflavonoid-proteasome interactions, offering valuable insights for developing novel anticancer proteasome inhibitors.

Keywords: Anticancer, *Araucaria*, Biflavonoid, Molecular docking, Molecular dynamics, MM/GBSA, Proteasome inhibitors, 3D-QSAR.

1. Introduction

Cancer is one of the leading causes of morbidity and mortality in the world, with more than 9.6 million deaths each year and continuing to increase. Cancer is caused by uncontrolled growth of abnormal cells, accompanied by metastasis to other body parts. Various studies are being conducted to develop more effective anticancer drugs, focusing on specific targets. Proteasomes, multi-catalytic protease complexes that play a role in the degradation of 80% of intracellular proteins through three catalytic sites with different substrate

[1] Corresponding Authors

e-mail: purwantiningsih@apps.ipb.ac.id

specificities (β1, β2, and β5), are one of the drug targets in cancer treatment that continue to attract attention [1]. Several pharmacological strategies have been carried out over the past few decades to design drugs that specifically target the proteolytic activity of the proteasome [2]. Proteasomes have significantly higher activity in cancer cells than normal cells, indicating their crucial role in cancer cell growth and survival [3-4]. Inhibition of proteasome activity is a promising approach in cancer therapy because it can inhibit cancer cell growth by preventing the degradation of proteins that regulate the cell cycle and cellular functions and disrupting NF-κB signalling, which plays a vital role in cancer development [5-6]. Therefore, developing proteasome inhibitors has become a significant focus in searching for more effective anticancer therapies. Some proteasome inhibitors currently used for cancer treatment include bortezomib, ixazomib, and marizomib. Most proteasome inhibitors, both those that have been approved and those in clinical trials, are peptide-based inhibitors and inhibit the β5 site of the proteasome. However, peptide-based inhibitors have several weaknesses and limitations, especially related to severe side effects and drug resistance caused by strong covalent and irreversible interactions with the proteasome catalytic site. Therefore, developing non-peptide and reversible (non-covalent) proteasome inhibitors that lead to a natural approach is urgent to overcome the weaknesses and limitations of peptide-based proteasome inhibitors.

Biflavonoids, a subclass of flavonoids, have demonstrated significant potential as non-peptide and non-covalent proteasome inhibitors. Several biflavonoid compounds have been found to exhibit inhibitory activity against the 20S proteasome, thereby presenting potential as anticancer agents. For example, moreloflavone and talbotaflavone, isolated from *Garcinia lateriflora*, have been reported to strongly inhibit the proteasome with IC_{50} values of 1.3 μM and 4.4 μM, respectively [7]. Additionally, biflavonoid compounds isolated from *Ginkgo biloba*, isoginkgetin, directly inhibit the 20S proteasome through three catalytic site activities at once, caspase-like (β1), trypsin-like (β2), and chymotrypsin-like activity (β5), and interfere with NF-κB signalling with an IC_{50} value of 11.2 μM [8]. Isoginkgetin effectively kills multiple myeloma cell lines (blood cancer) in vitro, even showing a higher and faster level of cell death induction (apoptosis) than peptide-based inhibitors (bortezomib) [8]. This compound has also demonstrated inhibitory activity against MCF-7 and HeLa cell lines, with IC_{50} values of 2.14 μM and 11.03 μM, respectively [9]. This discovery paves the way to explore the potential of biflavonoid compounds as alternative better anticancer drug candidates. However, to date, very little literature discusses the possibility of biflavonoids as anticancer drugs based on their ability to inhibit the 20S proteasome, especially regarding the structural features that accommodate their activity.

Biflavonoids are secondary metabolites found in various types of plants, showing promising potential in drug development. One of the rich sources of biflavonoids is the genus *Araucaria*; however, the potential of biflavonoids from this genus still needs to be explored. Of the 19 species in the genus *Araucaria*, nine are known to contain biflavonoid compounds. This study focuses on evaluating the potential of 22 biflavonoid compounds isolated from the genus *Araucaria* as proteasome inhibitors using an in silico approach.

Advancements in computer-aided drug design (CADD) have significantly improved the efficiency of drug discovery, reducing the time and cost required for experimental screening. One such computational technique, quantitative structure-activity relationship (QSAR) modelling, allows for the identification of structural features that influence biological activity. The QSAR approach allows researchers to gain deeper insights into the structural parameters that influence the proteasome inhibitory activity of biflavonoids. This study employs the 3D-QSAR technique, a significant advancement over the traditional 2D-QSAR approach, which is limited to the consideration of physicochemical parameters such as hydrophobicity, electrostatic and steric effects (classical Hans analysis) [10]. 3D-QSAR comprehensively accounts for all atomic properties within a compound, particularly those associated with the spatial representation of a molecule. This capability is enabled by a range of descriptor methods, including CoMFA (Comparative Molecular Field Analysis), CoMSIA (Comparative Molecular Similarity Indices Analysis), CoMBINE

(Comparative Binding Energy Analysis), COMMA (Comparative Molecular Moment Analysis) [10]. The 3D-QSAR approach offers a more rational and in-depth method to identify and predict the activity of potential proteasome 20S inhibitors. The potential of this technique to provide a comprehensive understanding of the structural parameters that influence the proteasome inhibitory activity of biflavonoids instills confidence in the robustness of our research methods. In tandem with the advancement of cheminformatics technology, QSAR has undergone a progressive evolution, leading to the development of increasingly sophisticated methods, such as 4D-QSAR, 5D-QSAR, 6D-QSAR, 7D-QSAR and Holo-QSAR. A thorough examination of the diverse QSAR methodologies has been comprehensively documented in numerous review articles [10-12].

The objective of this study is to examine the capacity of biflavonoids to function as chymotrypsin-like (β5) proteasome inhibitors, taking into account their encouraging non-covalent interactions with the catalytic site of the 20S proteasome. To this end, a 3D-QSAR model was developed, employing a curated dataset of flavonoid derivatives for which proteasome inhibitory activities have been empirically documented. This model is designed to serve as a predictive tool for identifying structural features that contribute to enhanced activity. Utilizing the developed 3D-QSAR model, we conducted a virtual screening of 22 biflavonoids from the genus *Araucaria*, identifying the five most promising candidates with the highest predicted inhibitory potency. To further evaluate their binding affinity and molecular interactions, molecular docking simulations were performed. The docking analysis yielded insights into the key binding interactions within the proteasome's β5 subunit and served as an essential preparatory step for the subsequent molecular dynamics (MD) simulations. The MD simulations were conducted to assess the stability of the ligand-proteasome complexes over time. To validate the predicted top-ranked biflavonoids from the 3D-QSAR model, structural fluctuations, protein-ligand interactions, and dynamic binding stability were analyzed to ascertain whether they retain their favourable binding conformations under physiological conditions. Furthermore, a series of molecular mechanics-generalized Born and surface area solvation (MM/GBSA) calculations were conducted to assess the free binding energy of these biflavonoids. This approach served to provide additional verification that the ranking observed in the 3D-QSAR model is consistent with their dynamic binding affinity. This study offers a comprehensive evaluation of biflavonoid-proteasome interactions by integrating 3D-QSAR modelling, molecular docking, MD simulations, and MM/GBSA calculations. The findings contribute to a deeper understanding of the structural requirements for effective proteasome inhibition and highlight biflavonoids as promising candidates for future anticancer drug development.

2. Computational Method

2.1 Dataset

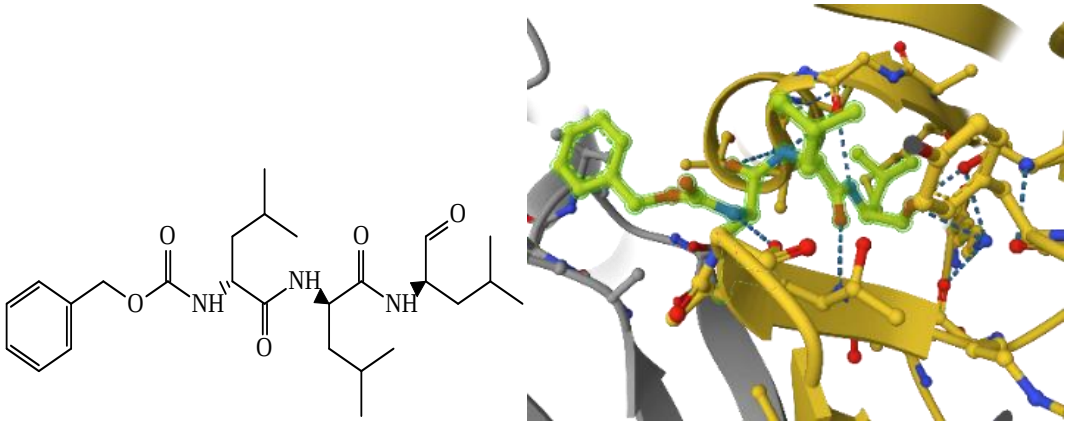
A total of 62 compound and activity data used as datasets in this study were obtained through literature studies [7, 8, 13-28]. The inhibitory activity data in IC_{50} were converted to pIC_{50} (decadic logarithm of IC_{50}). The pIC_{50} value was then used as the dependent variable for modelling. The dataset was then divided into a training set (50 compounds, 80%) and a test set (12 compounds, 20%).

2.2 3D-QSAR Modelling Using the PLS Algorithm on the OPEN3DQSAR Tool

Molecular alignment was performed in the Open3DAlign program using atom-based and pharmacophore alignment methods by setting the MG132 structure as an alignment template [29]. MG132 is a potent proteasome inhibitor with an IC_{50} value of 0.04 μM. The MG132 structure used as the alignment format is the conformational form of MG132 when bound to the β5 site of the proteasome **(Figure 1).**

In the modelling process employing Open3DQSAR, the molecular descriptor utilized is known as Molecular Interaction Fields (MIFs), a three-dimensional (3D) potential map that describes the interaction energy formed around the molecule [30-31]. The MIF descriptor is derived from the calculation of the interaction energy between each molecule aligned with the probe atom placed at grid points around the target molecule in a 3D lattice [30]. This study calculated the steric (van der Waals) and electrostatic (Coulombic) interaction energies using a sp^3 carbon atom probe

with a charge of +1. The energy was calculated using the MMFF94 force field in a 3D cubic lattice with a grid spacing of 1 Å. After the MIFs descriptor is calculated, the following standard data pre-treatment is performed to exclude less informative variables [31].



(A) (B)

Figure 1. 2D structure of MG132 (A) dan conformation when bound to the β5 site of the proteasome (B)

- Exclusion of grid points that exceed the cut-off in a specific MIF (excludes grid points that are very close to the atomic nucleus because they can produce very high steric energy values. The steric energy threshold is set at 10^4 kcal/mol).
- Max and min cut-off (set grid points that lie above or below these thresholds, respectively, to a user-defined maximum/minimum threshold value. In this case, the threshold was set from -30 to 30 kcal/mol to avoid extreme values in the data that could cause the model to be highly biased).
- Zeroing (sets to zero grid values which are close to zero).
- Standard deviation cut-off (removes variables that have a standard deviation between objects that is lower than a user-defined threshold, to improve the signal-to-noise ratio. The standard deviation threshold was set to 0.1).
- N-level variable elimination (removes variables that have only a few different values across the different objects to prevent them from biasing the model). Also, the following clustering and variable selection procedures are implemented in Open3DQSAR to enhance the predictive performance of the 3D-QSAR model.
- Smart Region Definition (SRD), as introduced by Pastor and co-workers [32], organizes variables based on their original spatial distribution in three-dimensional space. This approach minimizes redundancy by clustering nearby descriptors that convey essentially the same type of information.
- Fractional Factorial Design (FFD) variable selection, as initially proposed by Baroni and co-workers [33] is designed to identify variables with the greatest impact on predictive performance. This method can be applied to individual variables or to groups of variables identified through a prior SRD analysis.

The regression analysis employed in Open3DQSAR is Partial Least Squares (PLS), a methodology that has been selected in this study due to its ability to handle multicollinearity and high-dimensional datasets [34], which are commonplace in QSAR modelling. In contrast to Multiple Linear Regression (MLR), a conventional linear regression method that directly models the relationship between multiple independent variables and a dependent variable, PLS extracts principal components (PC) that maximize the variance between molecular descriptors and biological activity, thereby effectively mitigating

multicollinearity issues. Additionally, while MLR becomes impractical when the number of molecular descriptors exceeds the number of compounds, PLS overcomes this limitation by employing a projection-based approach. This characteristic renders PLS particularly well-suited for 3D-QSAR studies, ensuring the development of a more stable, generalizable, and predictive model.

The determination of the optimal number of principal components (PC) in constructing the PLS model is achieved through the use of internal and external validation. The external validation is the primary focus of this study, as it utilizes all available biflavonoid compound data with IC_{50} values (only three compounds) as a test set. This approach ensures the model's capacity to accurately predict the activity of biflavonoid compounds.

2.3 3D-QSAR Modelling Using Machine Learning Algorithms on the Orange Device

In the framework of machine learning-based modelling, molecular alignment was executed through the use of Discovery Studio Visualizer software, employing the identical compound template as PLS modelling. This alignment process incorporates steric and electrostatic factors, allocating a proportion of 50% to each factor (by default). Concurrently, three-dimensional (3D) molecular descriptors were computed utilizing the Mordred program. Mordred is a descriptor calculation program that is notable for its accessibility, ease of installation, expeditious calculation performance, and extensive flexibility of use [35]. In this study, we calculated 215 3D descriptors, which were categorized into several modules, including CPSA (Charged Partial Surface Area), Geometrical Index, Gravitational Index, MoRSE (Molecular Representation of Structures based on Electron diffraction), Moment of Inertia, and PBF (Polarizability Based Field). Each descriptor plays an indispensable role in the description of complex molecular characteristics and the prediction of biological activity.

Various machine learning algorithms were used in the modelling, including neural network, gradient boosting, support vector machine (SVM), random forest, decision tree, AdaBoost, and K-nearest neighbours (kNN). The modelling process used Orange software, an open-source data analysis and visualization tool [36]. The best model was selected based on internal, external, and cross-validation. Cross-validation was performed using the leave-one-out (LOO) method and 10-fold cross-validation. In addition, other parameters such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE) and standardized residual distribution were also considered. The selected model was then used to predict the IC_{50} of 22 biflavonoid compounds successfully isolated from the genus *Araucaria* (**Table 1**) [37-38].

Table 1. Biflavonoids from the genus *Araucaria* [37-38]

Structure	Biflavonoid Compounds	Substituents			
		R1 (7)	R2 (4')	R3 (7")	R4 (4"')
	(1) 7-*O*-methylagathisflavone	$-CH_3$	-H	-H	-H
	(2) 7"-*O*-methylagathisflavone	-H	-H	$-CH_3$	-H
	(3) 7,7"-di-*O*-methylagathisflavone	$-CH_3$	-H	$-CH_3$	-H
	(4) 7,4"'-di-*O*-methylagathisflavone	$-CH_3$	-H	-H	$-CH_3$
	(5) 7,7", 4"'-tri-*O*-methylagathisflavone	$-CH_3$	-H	$-CH_3$	$-CH_3$

Table 1. Biflavonoids from the genus *Araucaria* [37-38]

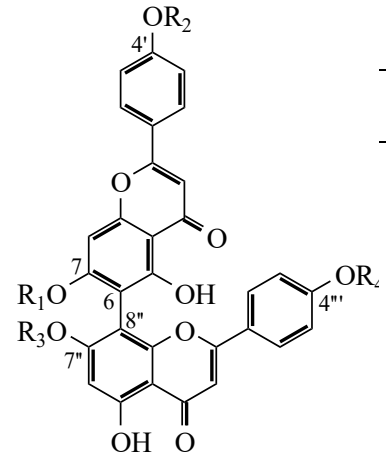
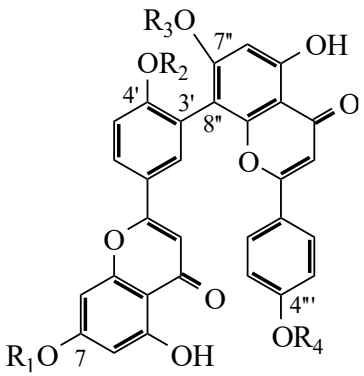
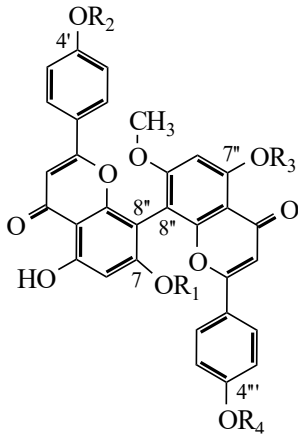
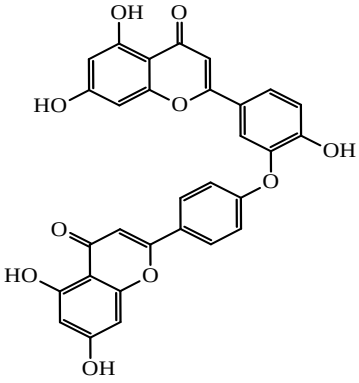
Structure	Biflavonoid Compounds	Substituents			
		R1 (7)	R2 (4')	R3 (7")	R4 (4"')
	**(6)** 7,4',7"-tri-*O*-methylagathisflavone	-CH_3	-CH_3	-CH_3	-H
	(7) 7,4',7",4"'-tetra-*O*-methylagathisflavone	-CH_3	-CH_3	-CH_3	-CH_3
	(8) 4′,7"-di-*O*-methylagathisflavone	-H	-CH_3	-CH_3	-H
	**(9)** 7,4′,7″-tri-*O*-methylamentoflavone	-CH_3	-CH_3	-CH_3	-H
	(10) 7,4′,4‴-tri-*O*-methylamentoflavone	-CH_3	-CH_3	-H	-CH_3
	(11) 7″-*O*-methylamentoflavone	-H	-H	-CH_3	-H
	(12) 7,7″-di-*O* methylamentoflavone	-CH_3	-H	-CH_3	-H
	(13) 7,4',7″,4‴-tetra-*O*-methylamentoflavone	-CH_3	-CH_3	-CH_3	-CH_3
	(14) 7,4'-di-*O*-methylamentoflavone	-CH_3	-CH_3	-H	-H
	**(15)** 7-*O*-methylcupressuflavone	-CH_3	-H	-H	-H
	(16) 7,7″-di-*O*-methylcupressuflavone	-CH_3	-H	-CH_3	-H
	(17) 7,4′,7″-tri-*O*-methylcupressuflavone	-CH_3	-CH_3	-CH_3	-H
	(18) 4′,4‴-di-*O*-methylcupressuflavone	-H	-CH_3	-H	-CH_3
	(19) 7,4′,7″,4‴-tetra-*O*-methylcupressuflavone	-CH_3	-CH_3	-CH_3	-CH_3
	(20) 7,7″,4‴-tri-*O*-methylcupressuflavone	-CH_3	-H	-CH_3	-CH_3
	(21) 7,4‴-di-*O*-methylcupressuflavone	-CH_3	-H	-H	-CH_3

Table 1. Biflavonoids from the genus *Araucaria* [37-38]

Structure	Biflavonoid Compounds	Substituents			
		R1 (7)	R2 (4')	R3 (7")	R4 (4"')
	**(22)** Ochnaflavon				

2.4 Molecular Docking

Molecular docking simulations were conducted using AutoDock Vina v1.2.6 to obtain the initial binding conformations of selected biflavonoids within the chymotrypsin-like (β5) active site of the 20S proteasome. The crystal structure of the 20S proteasome complexed with MG132 (PDB ID: 8CVR) was retrieved from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). Protein preparation involved removing water molecules and non-essential ions, followed by the addition of polar hydrogens and Gasteiger charges using AutoDockTools [39]. The binding site was defined based on the co-crystallized ligand MG132 (LDZ). Ligand structures, including five selected biflavonoids, were optimized using the MMFF94 force field and converted into PDBQT format to allow flexible torsion handling. It is important to emphasize that docking was performed primarily as a preparatory step for molecular dynamics (MD) simulations rather than as the main research objective. The top-ranked docking conformations were selected based on their binding affinity (kcal/mol) and stability, serving as the starting structures for subsequent MD simulations.

2.5 Molecular Dynamics Simulation

Molecular dynamics (MD) simulations were performed to assess the binding stability and dynamic behaviour of the biflavonoid-proteasome complexes. The simulations were conducted using Google Colab Pro+, which provides a cloud-based computational environment equipped with the necessary libraries and packages for MD simulations, including AmberTools, OpenMM, PyTraj, py3Dmol, NumPy, ProLIF, Matplotlib and Anaconda. These tools facilitated system preparation, trajectory analysis, and visualization, ensuring an efficient and reproducible simulation workflow.

The ff19SB force field was used for the protein, while the GAFF2 force field was applied to generate ligand topology. The system was solvated using the TIP3P water model in a 12 Å periodic box, with 0.15 M NaCl. Energy minimization was conducted with 20,000 steps before equilibration. The equilibration phase was performed for 5 ns with an integration step of 2 fs, maintaining a temperature of 298 K and pressure of 1 bar with a force constant of 700 kJ/mol. The production MD simulations were carried out for 100 ns under the same conditions. Post-simulation analysis included root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration (Rg) to evaluate system stability. The MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) method was used to estimate the binding free energy of the protein-ligand complexes over the final 10 ns of the molecular dynamics trajectory.

3. Results and discussion

3.1 3D-QSAR Model with PLS Algorithm

The 3D-QSAR model with the Comparative Molecular Field Analysis (CoMFA) approach was developed using 50 compounds as training sets and 12 test sets on the Open3DQSAR tool. CoMFA uses MIFs as molecular descriptors built using steric probes and electrostatic probes. The

contributions of steric and electrostatic fields in the model were 14.4% and 85.6%, respectively.

Principal component (PC) extraction and Partial Least Squares (PLS) modelling were performed to reduce the complexity of the variables in this study. Initially, 15 PCs were calculated. The results of PC extraction, as shown in **Figure 2**, show that an increase in the number of extracted PCs correlates with an increase in the variance of the independent and dependent variables that the PLS model can explain. In addition, an increase in the number of PCs is also followed by the rise in the coefficient of determination (R^2) from internal validation, indicating a direct relationship between the number of PCs and the model's ability to explain data variability. However, when the number of PCs used becomes too large, there is a negative effect on the predictive ability of the PLS model, indicated by the decrease in the coefficient of determination of external validation (R^2_{pred}) as the number of PCs used increases. This phenomenon indicates that although adding PCs can increase the explanation of variance in independent variables and increase R^2, it can also reduce the model's ability to predict pIC_{50} values of new compounds not included in the training data. This decrease in predictive ability indicates overfitting in the PLS model. Overfitting occurs when the model fits the training data well but fails to generalize to new data (test set). An excessive number of PCs causes the model to capture noise or random fluctuations in the training data as part of the pattern, making the model appear more accurate than it is. Therefore, finding the optimal number of PCs to balance the model's ability to explain the training data and its reliability in predicting new data is crucial. The PLS model with 7 PCs was chosen based on the evaluation because it produced the most satisfactory external and internal validation determination coefficients, 0.75 and 0.98, respectively (**Figure 3A**). The model is also evaluated based on the distribution of its standardized residuals, as shown in **Figure 3B**. The residuals of the model are randomly and normally distributed, indicating good model performance and not ignoring important interactions between the variables used in the model.

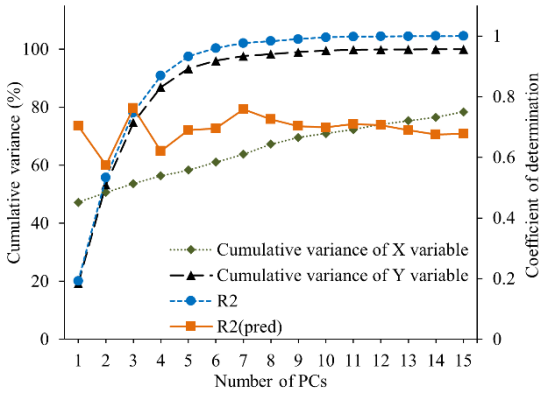


Figure 2. Cumulative variance curve of variables and coefficient of determination against the number of PCs

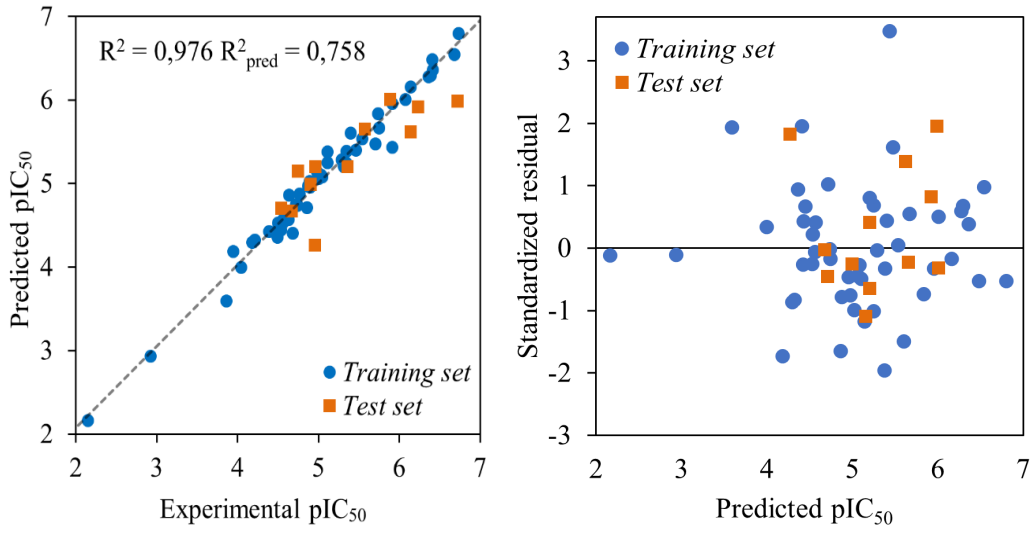
External validation, which is the main focus of this study, was carried out using all available IC50 values for biflavonoid compounds as a test set. This approach ensures that the model can accurately predict the activity of biflavonoid compounds, which is the primary goal of this research. The 3D quantitative structure-activity relationship (3D-QSAR) model, developed using Partial Least Squares (PLS), is represented as a 3D isocontour map of the PLS coefficients, as illustrated in Figure 4.

In this map, the green contour indicates regions where larger substituents (steric effects) enhance biological activity, while the yellow contour shows areas where larger substituents diminish biological activity. The blue contours highlight areas where positively charged substituents and hydrogen bond donors contribute to increased activity. Conversely, the red contour indicates that negatively charged substituents and hydrogen bond acceptors are positively correlated with biological activity. The isocontour map is further analyzed about three

biflavonoid compounds included in the test set: isoginkgetin, moreloflavone, and talbotaflavone, as shown in **Figure 4.**

3.2 3D-QSAR Model with Machine Learning Algorithm

The statistical parameters of modelling results using several machine learning algorithms with 3D descriptors from Mordred as independent variables are presented in **Table 2**. Among the models built, the Support Vector Machine (SVM) model showed the best predictive performance based on the results of external validation and cross-validation (10-fold cross-validation and leave-one-out cross-validation), with R^2_{pred}, $Q^2_{10\text{-}fold}$, and Q^2_{loo} values of 0.890, 0.422, and 0.416, respectively (**Table 2**).



(A) (B)

Figure 3. The plot of the predicted pIC_{50} values of the PLS model against the experimental values (A) and Distribution of residuals from the predicted pIC_{50} PLS model (B)

The SVM model also showed excellent performance in internal validation with an R^2 value of 0.952 (**Figure 5A**). SVM is a prevalent machine learning technique because it handles high-dimensional data effectively and often produces higher predictive performance than other models [40-41]. In contrast, other models, such as gradient boosting, AdaBoost, and neural networks, showed signs of overfitting, with very high R^2 values but low R^2_{pred}, $Q^2_{10\text{-}fold}$, and Q^2_{loo} values (**Table 2**). External validation is the main parameter for selecting machine learning-based models for the same reasons as in PLS modelling.

The performance of the SVM model is evaluated based on various parameters. These include the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), with values of 0.194, 0.148, and 0.032 respectively. Lower values of these parameters indicate better predictive accuracy. Additionally, the distribution of residuals is important in evaluating the model's performance. The residual plot in **Figure 5B** shows a relatively random distribution with a shape close to normal, indicating that the SVM model can predict data accurately and without significant systematic tendencies.

As the best model, the SVM model was then used to predict the proteasome inhibitory activity of 22 biflavonoid compounds successfully isolated from the genus *Araucaria* (**Table 1**) [37-38]. The prediction results obtained from the SVM model are shown in **Table 3.** Compounds are classified as highly active proteasome inhibitors if IC_{50} <1 µM, active if 1 µM <IC_{50}<4 µM, and weak if IC_{50} >4 µM [42].

Based on the IC_{50} prediction results of biflavonoid compounds tested using the SVM model, we have identified five active compounds with the potential to be proteasome inhibitors. Each compound has IC_{50} prediction values below 4 µM (Table 3), indicating their promising potential. Compound **1**, in particular, has emerged as the most promising, with an IC_{50} value of 1.764 µM. The next best sequences were **15**, **22**, **11**, and **2**, with IC_{50} values

of 2.083; 2.207; 2.437; and 2.948 µM, respectively. These five best compounds exhibited a higher prevalence of dominant −OH substituents than other compounds with lower IC_{50} predictions (**Figure 6**).

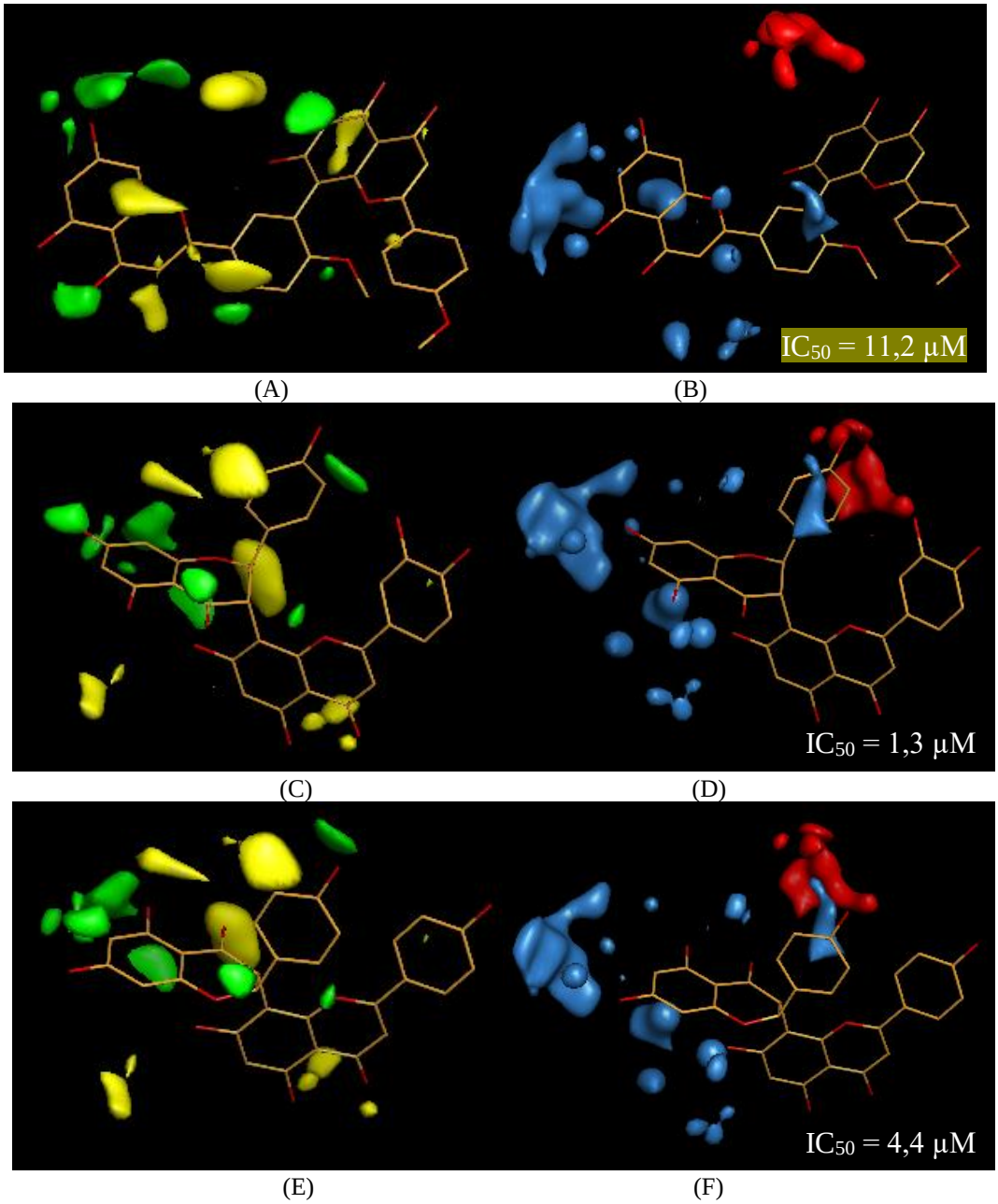
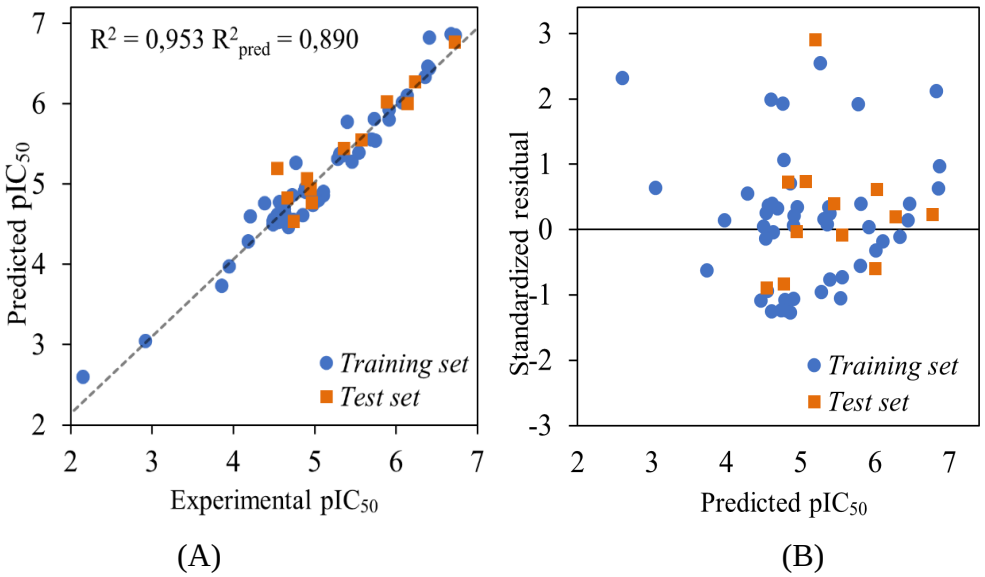


Figure 4. Isocontour of PLS coefficients. Isoginkgetin compounds with steric (A) and electrostatic (B) field isocontours, moreloflavon with steric (C) and electrostatic (D) field isocontours, talbotaflavone with steric (E) and electrostatic (F) field isocontours

Table 2. Statistical parameter values of machine learning-based models

Model	RMSE	MAE	MAPE	R^2	R^2_{pred}	$Q^2_{10\text{-}fold}$	Q^2_{loo}
SVM	0,194	0,148	0,032	0,952	0,890	0,422	0,416
kNN	0,601	0,419	0,103	0,542	0,718	0,321	0,310
Gradient Boosting	0,002	0,002	0,000	1,000	0,694	0,062	0,110
Random Forest	0,315	0,217	0,052	0,874	0,454	0,069	0,191
AdaBoost	0,042	0,011	0,002	0,998	0,400	0,067	0,180

Tree	0,253	0,126	0,035	0,919	-0,684	-0,752	-0,610
Neural Network	0,079	0,048	0,010	0,992	-5,045	-3,407	-4,018



(A) (B)

Figure 5. The plot of the predicted pIC_{50} values of the PLS model against the experimental values (A) and distribution of residuals from the predicted pIC_{50} PLS model (B)

Table 3. IC_{50} predicted results of biflavonoid compounds

Biflavonoid Compounds	Predicted IC_{50} (μM)	IC_{50} against MCF-7 (μM) [9]	IC_{50} against HeLa (μM) [9]
7-*O*-methylagathisflavone (**1**)	1,764	-	-
7-*O*-methylcupressuflavone (**15**)	2,083	3,40	1,42
Ochnaflavon (**22**)	2,207	-	-
7″-*O*-methylamentoflavone (**11**)	2,437	-	-
7″-*O*-methylagathisflavone (**2**)	2,948	-	-
7,7″-di-*O*-methylcupressuflavone (**16**)	4,798	-	-
4′,4‴-di-*O*-methylcupressuflavone (**18**)	5,090	-	-
7,4‴-di-*O*-methylagathisflavone (**4**)	5,539	-	-
7,4‴-di-*O*-methylcupressuflavone (**21**)	6,441	11,54	35,59
7,4′-di-*O*-methylamentoflavone (**14**)	6,922	-	-
7,7″-di-*O*-methylamentoflavone (**12**)	8,521	-	-
7,7″-di-*O*-methylagathisflavone (**3**)	9,033	115,39	107,63
7,7″,4‴-tri-*O*-methylcupressuflavone (**20**)	10,458	-	-
4′,7″-di-*O*-methylagathisflavone (**5**)	10,629	-	-
7,4′,7″-tri-*O*-methylamentoflavone (**9**)	11,792	-	-
7,4′,7″-tri-*O*-methylcupressuflavone (**17**)	13,144	91,74	-
7,4′,7″-tri-*O*-methylagathisflavone (**6**)	13,396	-	-
7,7″,4‴-tri-*O*-methylagathisflavone (**7**)	16,802	314,44	337,05
7,4′,4‴-tri-*O*-methylamentoflavone (**10**)	25,002	-	-
7,4′,7″,4‴-tetra-*O*-methylcupressuflavone (**19**)	26,758	397,89	528,78
7,4′,7″,4‴-tetra-*O*-methylamentoflavone (**13**)	34,786	-	-
4′,4‴,7,7″-tetra-*O*-methylagathisflavone (**8**)	56,554	-	-

Four of the five compounds had only one $-OCH_3$ substituent, all located at position 7 or 7″, except for ochnaflavone, which had all its substituents as −OH groups. The trend indicates that the more −OH substituents (the fewer $-OCH_3$ substituents), the higher the proteasome inhibitory activity of the compound (**Figure 7**). This underscores the crucial role of the −OH group in enhancing proteasome inhibitory activity. These findings align with the results of the PLS model interpretation, which

suggests that substituents acting as hydrogen bond donors (such as −OH groups) have a more significant effect in increasing proteasome inhibitory activity than hydrogen bond acceptors and bulk substituents (such as $-OCH_3$ groups).

Furthermore, these results are consistent with the findings of Piwowar and coworkers [43], who demonstrated that methylated flavonoids do not act as effective proteasome inhibitors in both in vitro and in vivo systems and only weakly induce apoptosis. In contrast, non-methylated flavonoids effectively inhibit proteasome activity in HL60 cells, leading to the accumulation of proteasomal target proteins and the activation of caspase-dependent apoptosis. More specifically, their study showed that the failure of methylated flavonoids to induce apoptosis was due to their inability to inhibit the chymotrypsin-like activity of the cellular proteasome, the same target site investigated in this study.

The IC_{50} prediction results were also compared with the in vitro studies of compounds **3**, **7**, **15**, **17**, **19**, and **21** against MCF-7 and HeLa cells [9]. As a result, there was agreement in ranking the most active compounds even though the IC_{50} values were quite different (**Table 3**).

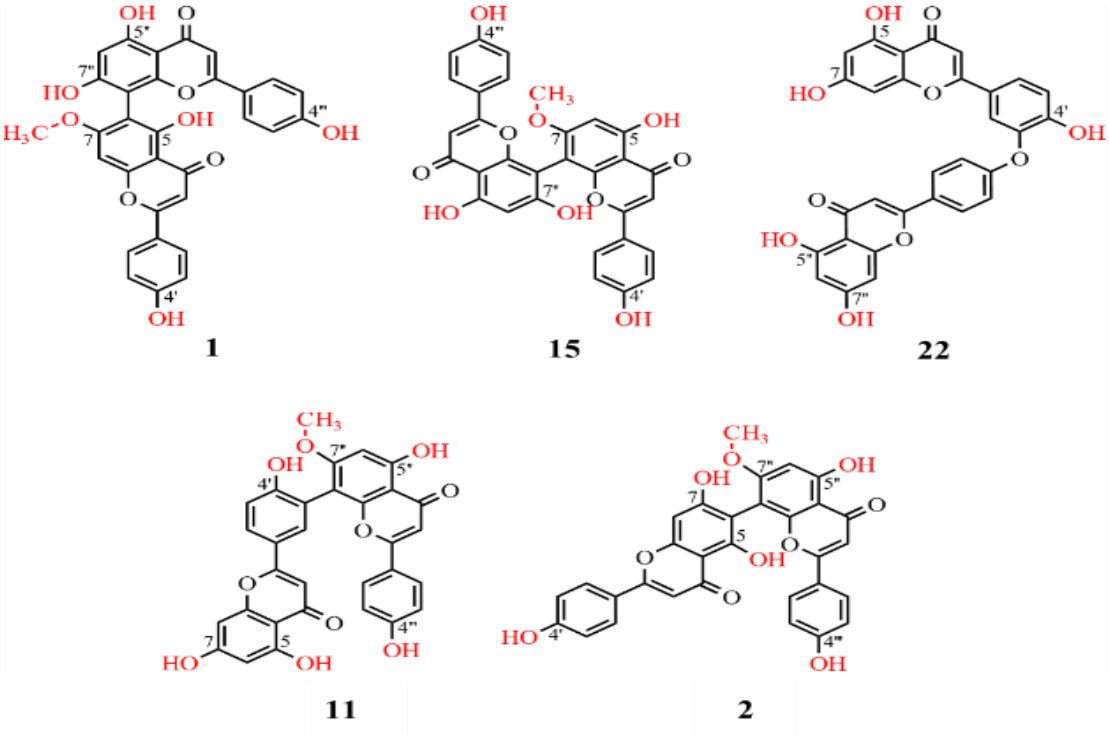


Figure 6. Structure of the 5 best compounds based on the SVM model prediction results

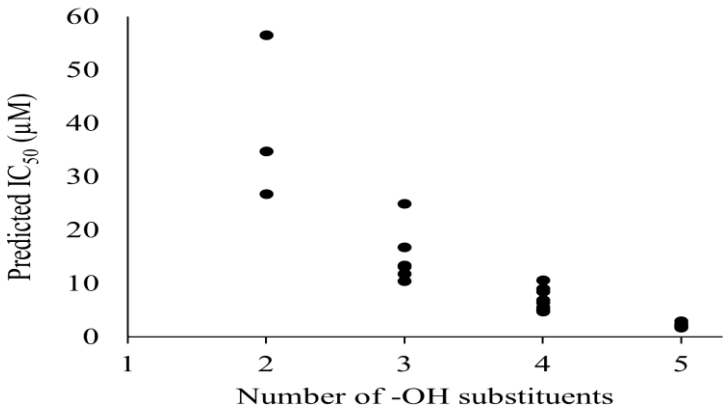


Figure 7. Structure of the 5 best compounds based on the SVM model prediction results.

Table 4. The docking-based binding affinity of the top five biflavonoid compounds

Biflavonoid Compounds	Binding Affinity (kcal/mol)
7-*O*-methylagathisflavone (**1**)	-9.319
7-*O*-methylcupressuflavone (**15**)	-9.041
Ochnaflavon (**22**)	-10.26
7″-*O*-methylamentoflavone (**11**)	-9.430
7″-*O*-methylagathisflavone (**2**)	-9.234

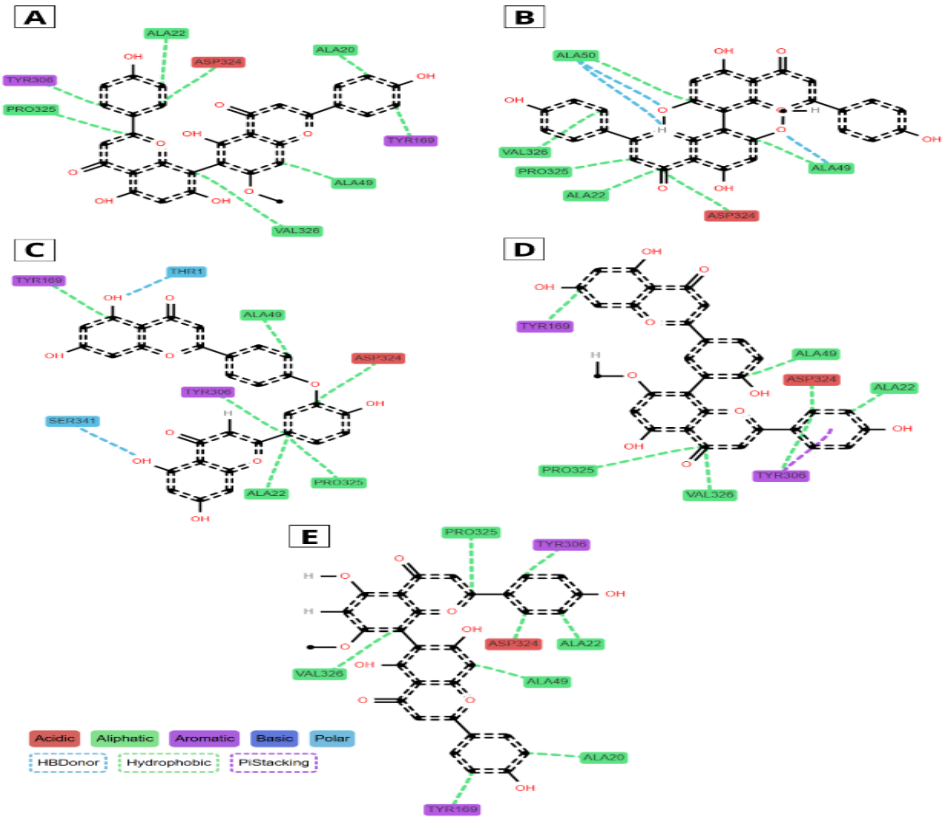


Figure 8. 2D view of the interactions of compounds **1** (A), **15** (B), **22** (C), **11** (D), and **2** (E) within the chymotrypsin-like (β5) active site of the 20S proteasome

3.3 The Analysis of Molecular Docking

Molecular docking was conducted to predict the binding interactions of the selected biflavonoids with the chymotrypsin-like (β5) active site of the 20S proteasome, providing an initial assessment of binding affinity and ligand orientation before molecular dynamics (MD) simulations. The results indicate that all five biflavonoids exhibit strong binding affinities, suggesting favourable interactions within the active site (**Table 4**).

Ochnaflavone (**22**) exhibited the strongest binding affinity of -10.26 kcal/mol, suggesting a robust interaction with key residues. The remaining compounds also demonstrated notable binding potential, exhibiting only minor variations in docking scores. These findings collectively indicate that the selected biflavonoids can effectively interact with the proteasome active site, supporting their potential role as competitive inhibitors.

A molecular docking interaction analysis was conducted to investigate the binding interactions between the selected biflavonoids and the chymotrypsin-like (β5) active site of the 20S proteasome. The docking interactions were then visualized in **Figure 8**, in which hydrogen bonds are represented as blue dashed lines, while hydrophobic interactions are depicted as green dashed lines. Hydrogen bonding plays a crucial role in stabilizing ligand binding by enhancing specificity and affinity, thereby strengthening proteasome inhibition. Meanwhile, hydrophobic interactions contribute significantly to ligand positioning within the active site, reinforcing

ligand-protein stability and facilitating favourable binding conformations [44].

The docking results revealed a consistent pattern of hydrophobic interactions among the five biflavonoid compounds, suggesting a shared binding mode within the β5 active site of the 20S proteasome. Compounds **1**, **2**, and **11** did not form hydrogen bonds, however, they engaged in multiple hydrophobic interactions with TYR306, TYR169, ALA20, ALA49, ASP324, VAL326, and PRO325 (**Figure 8A, 8D, 8E**). This finding suggests that these compounds rely predominantly on hydrophobic forces to stabilize their binding within the active site. In contrast, Compound **15** exhibited three hydrogen bonds with ALA50 and ALA49, in addition to hydrophobic interactions with VAL326, PRO325, ALA22, and ASP324, which may contribute to a stronger binding affinity (**Figure 8B**). Compound **22** formed two hydrogen bonds with THR1 and SER341 (**Figure 8C).** Of particular note is the significance of THR1, a pivotal catalytic residue in the proteolytic mechanism of the β5 active site, which plays a crucial role in peptide cleavage via nucleophilic attack [45]. Interaction with this residue is of particular relevance, as it may significantly impact proteasome inhibition. Furthermore, hydrophobic interactions with TYR306, ALA49, ASP324, and PRO325 appear to be a common feature across all five ligands, highlighting their role in stabilizing ligand binding within the β5 active site.

3.4 The Analysis of Molecular Dynamics Simulations

To further investigate the dynamic behaviour and stability of the five best-performing biflavonoid derivatives identified through 3D-QSAR modeling, molecular dynamics (MD) simulations were conducted. While molecular docking provides a static representation of ligand binding affinity, it does not account for the conformational flexibility of the protein-ligand complex under physiological conditions [46]. MD simulations, therefore, serve as an essential approach to evaluating the stability of these interactions over time.

In this study, MD simulations were performed for 100 ns using the docked conformations of each compound as the initial input structures. This approach ensures that the simulations reflect the most favourable binding orientations predicted by molecular docking. Key structural and energetic parameters, including root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), MM/GBSA binding free energy, and interaction energy, were systematically analyzed to assess the stability and binding characteristics of each complex [47].

Root mean square deviation (RMSD) is a fundamental parameter used to evaluate the overall stability of a protein-ligand complex during molecular dynamics simulations. It measures the conformational changes of the complex over time by calculating the deviation of atomic positions from the initial structure. A lower and more stable RMSD value indicates that the complex maintains its structural integrity throughout the simulation, suggesting a strong and stable binding interaction between the ligand and the protein. Conversely, significant fluctuations in RMSD may signify structural instability, conformational rearrangements, or potential dissociation of the ligand from the binding site.

The RMSD profiles of the five selected biflavonoid derivatives revealed distinct stability patterns (**Figure 9**). Compounds **1**, **2**, and **11** exhibited relatively stable RMSD values, maintaining fluctuations within the range of 2–3 Å. These complexes reached equilibrium after the first 10 ns and remained stable throughout the remaining simulation time. Such behaviour suggests that these compounds form stable interactions with the protein, experiencing only minimal conformational shifts. The observation of relatively low RMSD values indicates that the binding of these ligands does not induce significant perturbations in the protein structure, thereby reinforcing their potential as strong inhibitors. Among these three compounds, compound **1** displayed the highest stability, as evidenced by its minimal fluctuations, further emphasizing its strong affinity and compatibility with the binding site.

In contrast, compounds **15** and **22** exhibited more pronounced RMSD fluctuations, suggesting greater conformational flexibility or instability within the protein-ligand complex. Compound **22** fluctuated significantly around 6 Å and only began to stabilize after approximately 40 ns. This behaviour implies that the ligand required a longer equilibration period to achieve a relatively stable binding mode, possibly due to initial repositioning or

rearrangement within the binding pocket. Conversely, compound **15** exhibited a continuous increase in RMSD, reaching up to 8 Å by the conclusion of the 100 ns simulation, which illustrates that its interaction with the protein is unstable.

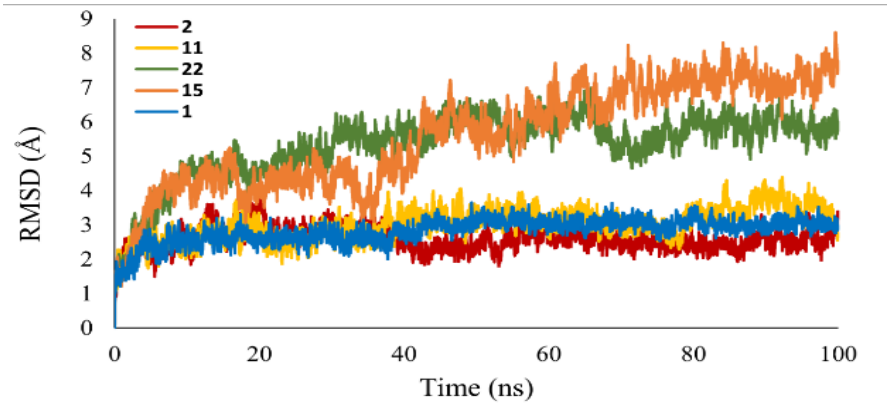


Figure 9. The RMSD values of ligands at the active site of 20S proteasome during 100 ns MD simulation

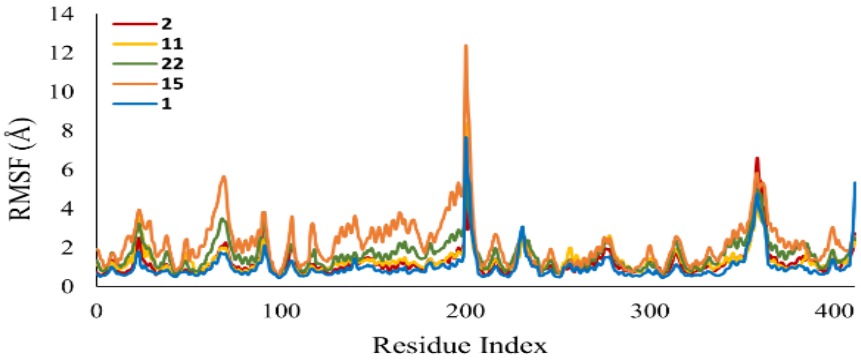


Figure 10. The RMSF values of five protein-ligand complexes

Root mean square fluctuation (RMSF) is a pivotal parameter employed to assess the flexibility of individual residues in a protein-ligand complex during molecular dynamics (MD) simulations. Lower RMSF values indicate residues that are more rigid and stable, whereas higher RMSF values suggest increased flexibility or potential structural rearrangements [48]. The RMSF plot, presented in **Figure 10**, illustrates the fluctuation of each residue over the 100 ns simulation period.

The majority of residues exhibit RMSF values below 4 Å, indicating that most of the protein remains structurally stable throughout the simulation. However, a few residues show significantly higher fluctuations. These peaks may correspond to loop regions, terminal ends, or flexible binding sites that contribute to the dynamic behaviour of the protein.

The analysis reveals that among the five compounds that were examined, compound **1** exhibits the lowest overall RMSF values, followed by compounds **2** and **11**. This observation suggests that these complexes maintain a relatively stable protein-ligand interaction with minimal disruption to the protein's structural integrity. In contrast, compounds **15** and **22** exhibit higher RMSF values across multiple residues, reflecting greater structural flexibility. This increased fluctuation may indicate a less stable binding mode or a tendency for the ligand to induce conformational changes in the protein, which could potentially impact binding affinity and functional stability.

The radius of gyration (Rg) is a measure of the spatial distribution of atoms concerning the centre of mass. It has emerged as a critical metric for assessing the compactness of proteins. Lower and more stable Rg values are indicative of well-folded

and structurally stable complexes, while significant fluctuations are suggestive of potential conformational changes or loss of structural integrity.

The results indicate that compounds **1** and **2** exhibit the most stable Rg values, fluctuating consistently around 22–23 Å throughout the simulation (**Figure 11**), suggesting that the protein-ligand complex maintains a compact structure. These results align with the findings from RMSD and RMSF, indicating that these ligands contribute to a more stable interaction within the binding pocket. In addition, compound **11** initially exhibited higher and more variable Rg values during the first 20 ns, suggesting an early conformational adjustment phase. However, after this initial fluctuation, the complex stabilized, displaying a pattern similar to compounds **1** and **2**.

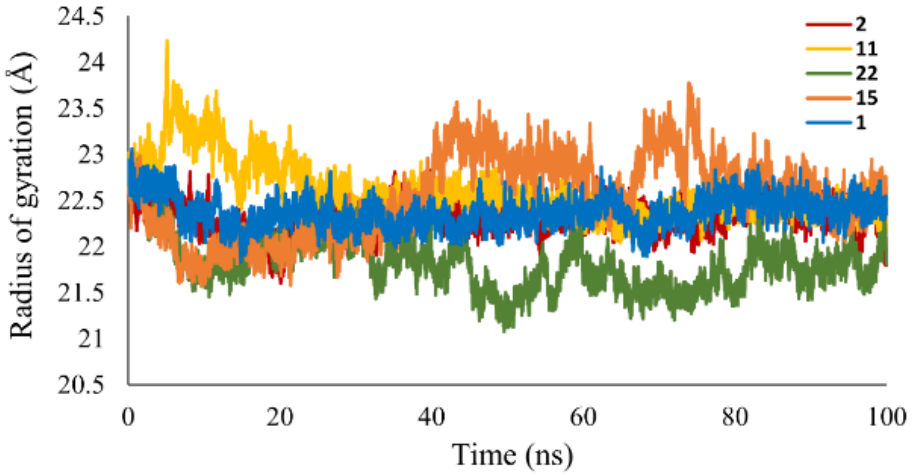


Figure 11. The radius of gyration values in five protein-ligan complexes

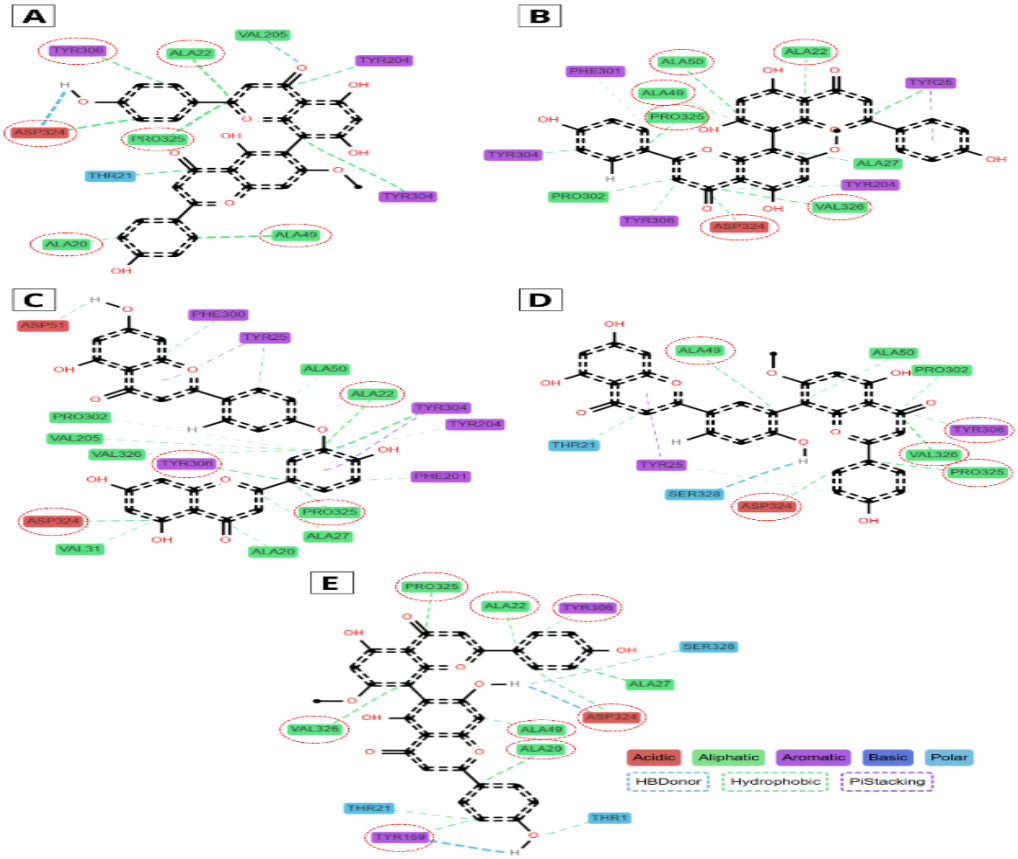


Figure 12. 2D view of the interactions of compounds **1** (A), **15** (B), **22** (C), **11** (D), and **2** (E) within the chymotrypsin-like (β5) active site of the 20S proteasome

This finding indicates that the ligand eventually adopts a more stable binding mode over time. Conversely, compounds **15** and **22** displayed greater fluctuations in Rg, reflecting a tendency toward structural instability. This observation is consistent with the higher RMSD and RMSF values, suggesting that these ligands may induce more substantial conformational changes in the protein or bind in a less stable manner.In order to gain deeper insights into the molecular interactions stabilizing the biflavonoid-proteasome complexes, an analysis of the interactions that occurred throughout the MD simulations was conducted. The 2D interaction diagram presented in **Figure 12** illustrates the interactions that appeared with a frequency exceeding 30% of the total simulation frames. Interactions above this threshold are considered relatively stable, indicating their potential significance in maintaining ligand binding within the active site.

The interaction diagram employs various dashed lines to denote distinct interaction types, blue dashed lines indicate hydrogen bonds, green dashed lines represent hydrophobic interactions, and purple dashed lines denote π-π stacking interactions. The thickness of these lines is proportional to the frequency of the interactions, with thicker lines signifying more persistent and stable interactions over the simulation period. This analysis enables a direct comparison between the molecular docking results and the MD-derived interaction patterns, providing a critical evaluation of whether the initially predicted docking interactions remain stable under dynamic conditions. Residues highlighted in red correspond to those that were also involved in key interactions observed during molecular docking, reinforcing their importance in ligand binding.

In general, the MD simulations revealed a greater number of interactions than those initially observed in molecular docking. Notably, several significant interactions, including hydrogen bonds and π-π stacking interactions, were formed during MD but were not predicted in docking calculations. This underscores the importance of incorporating protein flexibility in evaluating binding stability. In contrast, certain hydrogen bonds initially identified in docking, particularly for compounds **15** (**Figure 8B**) and **22** (**Figure 8C**), were absent in the MD interaction analysis, indicating that these interactions were not stable over time. This suggests that some hydrogen bonds predicted in docking might be transient rather than persistent, emphasizing the necessity of MD simulations in validating ligand stability.

The present comparative analysis between docking and MD interactions provides crucial insights into the binding stability and adaptability of biflavonoid inhibitors within the proteasome active site. The findings reinforce the role of MD simulations in refining docking predictions and identifying key interactions essential for ligand stabilization, thereby contributing to a more comprehensive evaluation of potential proteasome inhibitors.

Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) is a widely used computational approach for estimating the binding free energy of protein-ligand complexes in molecular dynamics simulations. Unlike molecular docking, which provides a static representation of binding affinity, MM/GBSA accounts for the solvent effects and dynamic nature of ligand binding by averaging energy contributions over multiple simulation frames [49]. The binding free energy (ΔG_{bind}) calculated using MM/GBSA is derived from the difference between the total energy of the complex and the sum of the energies of the individual protein and ligand components. A more negative ΔG_{bind} value indicates a stronger and more stable interaction. In addition to MM/GBSA, non-bonded interaction energy ($\Delta G_{non\text{-}bond}$) provides further insights into the stability and strength of ligand binding. Non-bonded interaction energy is chiefly constituted of van der Waals and electrostatic forces, which influence the extent to which a ligand remains associated with its target throughout the simulation. While MM/GBSA evaluates the overall thermodynamic favorability of binding, non-bonded interaction energy directly measures the physical interactions between the ligand and the protein. By analyzing both MM/GBSA binding free energy and interaction energy, we can better understand the relative stability and binding strength of the selected biflavonoid compounds within the proteasome active site. These results complement molecular docking scores and MD-based stability analyses (RMSD, RMSF, and Rg), offering a comprehensive evaluation of ligand-protein interactions.

The MM/GBSA binding free energy (ΔG_{bind}) and non-bonded interaction energy ($\Delta G_{non\text{-}bond}$) were evaluated for the five top-ranked compounds based on the predicted IC_{50} from the 3D-QSAR/SVM model (**Table 5**). The comparison between the ranking obtained from the MM/GBSA analysis and the initial IC_{50}-based ranking indicates an overall similarity, with a notable exception for compound **22** (Ochnaflavone).

Compound **22**, which ranked third in the 3D-QSAR-based IC_{50} predictions, exhibited the most favourable binding free energy of -34.9873 kcal/mol and the highest non-bonded interaction energy of -60.20 kcal/mol. This suggests that **22** forms the strongest interaction with the proteasome active site among the tested compounds. However, despite its favourable energetics, previous discussions on MD-derived parameters, including RMSD, RMSF, and radius of gyration (Rg), indicate that the complex formed by **22** is relatively unstable. This discrepancy highlights that strong binding energy alone does not necessarily guarantee structural stability during MD simulations.

Table 5. The binding free energy (MM/GBSA) and non-bonded interaction energy of biflavonoid-proteasome complexes

Biflavonoid Compounds	Predicted IC_{50} of 3D-QSAR model (μM)	MM/GBSA ΔG_{bind} (kcal/mol)	Interaction Energy $\Delta G_{non\text{-}bond}$ (kcal/mol)
7-*O*-methylagathisflavone (**1**)	1,764	-31.4897	-58.32
7-*O*-methylcupressuflavone (**15**)	2,083	-27.4875	-45.87
Ochnaflavon (**22**)	2,207	-34.9873	-60.20
7″-*O*-methylamentoflavone (**11**)	2,437	-24.8338	-54.00
7″-*O*-methylagathisflavone (**2**)	2,948	-20.9061	-46.32

Interestingly, structural analysis of these biflavonoids suggests a possible explanation for the unique behaviour of **22**. Unlike the other four compounds, which possess a direct C-C linkage between their flavonoid monomers, **22** is characterized by a C-O-C ether linkage. This difference in connectivity may contribute to greater conformational flexibility, potentially reducing the overall stability of its complex with the proteasome. Additionally, among the five candidate compounds, only **22** has hydroxyl (−OH) groups as its sole substituents, whereas the others all contain at least one methoxy ($-OCH_3$) group. The presence of multiple hydroxyl groups could enhance hydrogen bonding interactions but might also introduce additional solvent interactions that influence the ligand's dynamic stability within the binding pocket.

Overall, while **22** exhibits the most favourable binding energetics, its dynamic instability observed in MD simulations suggests that structural factors such as linkage type and substituent effects play a crucial role in modulating ligand behaviour. These findings underscore the importance of integrating both energetic and dynamic stability assessments when evaluating potential proteasome inhibitors.

4. Conclusions

The 3D-QSAR model was successfully constructed using 62 compounds as the data set, with the PLS model indicating that electrostatic factors play a dominant role in the proteasome inhibitory activity of biflavonoid compounds. Meanwhile, the machine learning-based 3D-QSAR model identified the SVM model as the most predictive, allowing for the selection of five active biflavonoid compounds from the genus *Araucaria* as potential proteasome inhibitors: 7-*O*-methylagathisflavone (**1**), 7-*O*-methylcupressuflavone (**15**), Ochnaflavone (**22**), 7″-*O*-methylamentoflavone (**11**), 7″-*O*-methylagathisflavone (**2**). These compounds exhibited higher numbers of hydroxyl substituents, further confirming the PLS model's findings that electrostatic interactions, particularly hydrogen bond donors, significantly contribute to proteasome inhibition. To further evaluate their binding interactions and stability, molecular docking and molecular dynamics (MD) simulations were conducted. The molecular docking results indicated that all five compounds exhibited strong binding affinities with the chymotrypsin-like (β5) active site of the 20S proteasome, with Ochnaflavone (**22**) demonstrating the highest docking score. However, molecular dynamics simulations provided a more comprehensive assessment of the stability of ligand-protein

complexes. The results demonstrated that compounds **1**, **2**, and **11** formed stable complexes with the proteasome active site, as evidenced by their low RMSD fluctuations, minimal RMSF values, and stable radius of gyration (Rg) over 100 ns of MD simulation. Among these, **1** demonstrated the most robust stability. In contrast, **15** and **22** exhibited increased conformational flexibility, indicating that their interactions with the proteasome were less stable.

Furthermore, MM/GBSA binding free energy calculations validated the ranking observed in 3D-QSAR predictions, with **1** and **22** demonstrating the strongest binding free energies. However, the dynamic instability of **22**, as observed in MD simulations, suggests that its strong initial binding may not translate into sustained inhibition under physiological conditions. A subsequent analysis of protein-ligand interactions during docking and molecular dynamics simulations revealed that all five compounds consistently interacted with key residues within the β5 active site, primarily through hydrophobic interactions and hydrogen bonds. The 2D interaction diagrams indicated that some hydrogen bonds observed in docking, particularly in compounds **15** and **22**, were not retained in MD simulations, suggesting that these interactions were transient rather than stable. Conversely, several novel interactions emerged during MD, including additional hydrogen bonds and π-π stacking interactions, underscoring the significance of protein flexibility in ligand binding. These findings underscore the necessity of MD simulations for validating the stability of interactions over time while acknowledging the value of docking in providing preliminary insights into potential interactions.

ACKNOWLEDGEMENTS

The authors are grateful to the Indonesian Ministry of Education, Culture, Research, and Technology through IPB University under Project No. 102/E5/PG.02.00.PL/2023, dated June 19, 2023.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

References

[1] Y. Kim, E.K. Kim, Y. Chey, M.J. Song, H.H. Jang, Targeted protein degradation: principles and applications of the proteasome, Cells, 12 (2023) 1846.

[2] M. Benvenuto, S. Ciuffa, C. Focaccetti, D. Sbardella, S. Fazi, M. Scimeca, G.R. Tundo, G. Barillari, M. Segni, E. Bonanno, V. Manzari, A. Modesti, L. Masuelli, M. Coletta, R. Bei, Proteasome inhibition by bortezomib parallels a reduction in head and neck cancer cells growth, and an increase in tumor-infiltrating immune cells, Sci Rep, 11 (2021) 1–23.

[3] C.L. Soave, T. Guerin, J. Liu, Q.P. Dou, Targeting the ubiquitin-proteasome system for cancer treatment: discovering novel inhibitors from nature and drug repurposing, Cancer and Metastasis Reviews, 36 (2017) 717–736.

[4] Q.P. Dou, J.A. Zonder, Overview of proteasome inhibitor-based anti-cancer therapies: Perspective on bortezomib and second generation proteasome inhibitors versus future generation inhibitors of ubiquitin-proteasome system, Curr Cancer Drug Targets, 14 (2014) 517.

[5] N. Guo, Z. Peng, MG132, a proteasome inhibitor, induces apoptosis in tumour cells, Asia Pac J Clin Oncol, 9 (2013) 6–11.

[6] F. Obrist, G. Manic, G. Kroemer, I. Vitale, L. Galluzzi, Trial Watch: Proteasomal inhibitors for anticancer therapy, Mol Cell Oncol, 2 (2015).

[7] Y. Ren, D.D. Lantvit, E.J.C. De Blanco, L.B.S. Kardono, S. Riswan, H. Chai, C.E. Cottrell, N.R. Farnsworth, S.M. Swanson, Y. Ding, X.C. Li, J.P.J. Marais, D. Ferreira, A.D. Kinghorn, Proteasome-inhibitory and cytotoxic constituents of Garcinia lateriflora: absolute configuration of caged xanthones, Tetrahedron, 66 (2010) 5311–5320.

[8] J. Tsalikis, M. Abdel-Nour, A. Farahvash, M.T. Sorbara, S. Poon, D.J. Philpott, S.E. Girardin, Isoginkgetin, a natural biflavonoid proteasome inhibitor, sensitizes cancer cells to apoptosis via disruption of lysosomal homeostasis and impaired protein clearance, Mol Cell Biol, 39 (2019).

[9] P. Sugita, D.D. Agusta, H. Dianhar, I.H. Suparto, K. Kurniawanti, D.U.C. Rahayu, L. Irfana, The cytotoxicity and SAR analysis of biflavonoids isolated from *Araucaria hunsteinii* K Schum leaves against MCF-7

and HeLa cancer cells, J Appl Pharm Sci, 13 (2023) 199–209.

[10] M. Damale, S. Harke, F. Kalam Khan, D. Shinde, J. Sangshetti, Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review, Mini-Reviews in Medicinal Chemistry, 14 (2014) 35–55.

[11] A.K. Halder, A.S. Moura, M.N.D.S. Cordeiro, QSAR modelling: a therapeutic patent review 2010-present, Expert Opin Ther Pat, 28 (2018) 467–476.

[12] A. Tropsha, O. Isayev, A. Varnek, G. Schneider, A. Cherkasov, Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR, Nat Rev Drug Discov, 23 (2024) 141–155.

[13] S. Nam, D.M. Smith, Q.P. Dou, Ester bond-containing tea polyphenols potently inhibit proteasome activity in vitro and in vivo, Journal of Biological Chemistry, 276 (2001) 13322–13330.

[14] D.M. Smith, Z. Wang, A. Kazi, L.H. Li, T.H. Chan, Q.P. Dou, Synthetic analogues of green tea polyphenols as proteasome inhibitors, Molecular Medicine, 8 (2002) 382–392.

[15] D. Chen, K.G. Daniel, M.S. Chen, D.J. Kuhn, K.R. Landis-Piwowar, Q.P. Dou, Dietary flavonoids as proteasome inhibitors and apoptosis inducers in human leukaemia cells, Biochem Pharmacol, 69 (2005) 1421–1432.

[16] F. Pajonk, J. Scholber, B. Fiebich, Hypericin-an inhibitor of proteasome function, Cancer Chemother Pharmacol, 55 (2005) 439–446.

[17] S.B. Wan, K.R. Landis-Piwowar, D.J. Kuhn, D. Chen, Q.P. Dou, T.H. Chan, Structure-activity study of epi-gallocatechin gallate (EGCG) analogues as proteasome inhibitors, Bioorg Med Chem, 13 (2005) 2177–2185.

[18] V. Milacic, S. Banerjee, K.R. Landis-Piwowar, F.H. Sarkar, A.P.N. Majumdar, Q.P. Dou, Curcumin inhibits the proteasome activity in human colon cancer cells in vitro and in vivo, Cancer Res, 68 (2008) 7283–7292.

[19] T.L. Chang, Inhibitory effect of flavonoids on 26S proteasome activity, J Agric Food Chem, 57 (2009) 9706–9715.

[20] C. Huo, H. Yang, Q.C. Cui, Q.P. Dou, T.H. Chan, Proteasome inhibition in human breast cancer cells with high catechol-O-methyltransferase activity by green tea polyphenol EGCG analogues, Bioorg Med Chem, 18 (2010) 1252–1258.

[21] H. Huang, N. Liu, K. Zhao, C. Zhu, X. Lu, S. Li, W. Lian, P. Zhou, X. Dong, C. Zhao, H. Guo, C. Zhang, C. Yang, G. Wen, L. Lu, X. Li, L. Guan, C. Liu, X. Wang, Q.P. Dou, et al., Sanggenon C decreases tumour cell viability associated with proteasome inhibition, Front Biosci, 3 (2011) 1315–1325.

[22] L. Pan, S. Matthew, D.D. Lantvit, X. Zhang, T.N. Ninh, H. Chai, E.J. Carcache De Blanco, D.D. Soejarto, S.M. Swanson, A.D. Kinghorn, Bioassay-guided isolation of constituents of Piper sarmentosum using a mitochondrial transmembrane potential assay, J Nat Prod, 74 (2011) 2193–2199.

[23] S.H. Shim, 20S proteasome inhibitory activity of flavonoids isolated from Spatholobus suberectus, Phytotherapy Research, 25 (2011) 615–618.

[24] M. Piedfer, S. Bouchet, R. Tang, C. Billard, D. Dauzonne, B. Bauvois, p70S6 kinase is a target of the novel proteasome inhibitor 3,3'-diamino-4'-methoxyflavone during apoptosis in human myeloid tumour cells, Biochim Biophys Acta, 1833 (2013) 1316–1328.

[25] Y. He, J. Huang, P. Wang, X. Shen, S. Li, L. Yang, W. Liu, A. Suksamrarn, G. Zhang, F. Wang, Y. He, J. Huang, P. Wang, X. Shen, S. Li, L. Yang, W. Liu, A. Suksamrarn, G. Zhang, F. Wang, Emodin potentiates the antiproliferative effect of interferon α/β by activation of JAK/STAT pathway signalling through inhibition of the 26S proteasome, Oncotarget, 7 (2015) 4664–4679.

[26] S.B. Wan, D. Chen, Q.P. Dou, T.H. Chan, Study of the green tea polyphenols catechin-3-gallate (CG) and epicatechin-3-gallate (ECG) as proteasome inhibitors, Bioorg Med Chem, 12 (2004) 3521–3527.

[27] A. Dreiseitel, P. Schreier, A. Oehme, S. Locher, G. Rogler, H. Piberger, G. Hajak, P.G. Sand, Inhibition of proteasome activity by anthocyanins and anthocyanidins, Biochem Biophys Res Commun, 372 (2008) 57–61.

[28] T.L. Chang, H.Y. Ding, K.N. Teng, Y.C. Fang, 5,6,3',4'-Tetrahydroxy-7-methoxyflavone as a novel potential proteasome inhibitor, Planta Med, 76 (2010) 987–994.

[29] P. Tosco, T. Balle, F. Shiri, Open3DALIGN: an open-source software aimed at unsupervised ligand alignment, J Comput Aided Mol Des, 25 (2011) 777–783.

[30] D. Hayakawa, N. Sawada, Y. Watanabe, H. Gouda, A molecular interaction field describing nonconventional intermolecular interactions and its application to protein-ligand interaction prediction, J Mol Graph Model, 96 (2020) 107515.

[31] P. Tosco, T. Balle, Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields, J Mol Model, 17 (2011) 201–208.

[32] M. Pastor, G. Cruciani, S. Clementi, Smart Region Definition: A New Way To Improve the Predictive Ability and Interpretability of Three-Dimensional Quantitative Structure−Activity Relationships, J Med Chem, 40 (1997) 1455–1464.

[33] M. Baroni, G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, S. Clementi, Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems, Quantitative Structure-Activity Relationships, 12 (1993) 9–20.

[34] D. Chung, S. Keles, Sparse Partial Least Squares Classification for High Dimensional Data, Stat Appl Genet Mol Biol, 9 (2010) 17.

[35] H. Moriwaki, Y.S. Tian, N. Kawashita, T. Takagi, Mordred: a molecular descriptor calculator, J Cheminform, 10 (2018) 4.

[36] J. Demšar, T. Curk, A. Erjavec, Č. Group, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, B. Zupan, Orange: Data mining toolbox in python, Journal of Machine Learning Research, 14 (2013) 2349–2353.

[37] P. Sugita, S.D.P. Handayani, D.D. Agusta, L. Ambarsari, H. Dianhar, D.U.C. Rahayu, Combined in-silico and in-vitro approaches to evaluate the inhibitory potential of biflavonoids from araucaria plants against α-glucosidase as target protein, Rasayan Journal of Chemistry, 16 (2023) 361–375.

[38] V.S. Gontijo, M.H. dos Santos, C. Viegas Jr., Biological and chemical aspects of natural biflavonoids from plants: A brief review, Medicinal Chemistry, 17 (2016).

[39] M. V. Shapovalov, R.L. Dunbrack, A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions, Structure, 19 (2011) 844–858.

[40] W. Sun, C. Chang, Q. Long, Bayesian non-linear support vector machine for high-dimensional data with incorporation of graph information on features, Proceedings - 2019 IEEE International Conference on Big Data, (2019) 4874–4882.

[41] B. Pes, Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains, Neural Comput Appl, 32 (2020) 5951–5973.

[42] S. Sengupta, G. Mehta, Non-peptidic natural products as ubiquitin-proteasome inhibitors, Tetrahedron, 75 (2019) 817–853.

[43] K.R. Landis-Piwowar, V. Milacic, Q.P. Dou, Relationship between the methylation status of dietary flavonoids and their growth-inhibitory and apoptosis-inducing activities in human cancer cells, J Cell Biochem, 105 (2008) 514.

[44] A.K. Varma, R. Patil, S. Das, A. Stanley, L. Yadav, A. Sudhakar, Optimized Hydrophobic Interactions and Hydrogen Bonding at the Target-Ligand Interface Leads the Pathways of Drug-Designing, PLoS One, 5 (2010) e12029.

[45] N. Vigneron, V. Ferrari, V. Stroobant, J.A. Habib, B.J. Van Den Eynde, Peptide splicing by the proteasome, J Biol Chem, 292 (2017) 21170.

[46] I.A. Guedes, C.S. de Magalhães, L.E. Dardenne, Receptor–ligand molecular docking, Biophys Rev, 6 (2013) 75.

[47] A.S. Abouzied, S. Alqarni, K.M. Younes, S.M. Alanazi, D.M. Alrsheed, R.K. Alhathal, B. Huwaimel, A.M. Elkashlan, Structural and free energy landscape analysis for the discovery of antiviral compounds targeting the cap-binding domain of influenza polymerase PB2,

Scientific Reports 2024 14:1, 14 (2024) 1–14.

[48] A.A. Alzain, F.A. Elbadwi, T.H. Shoaib, A.E. Sherif, W. Osman, A. Ashour, G.A. Mohamed, S.R.M. Ibrahim, E.J. Roh, A.H.E. Hassan, Integrating computational methods guided the discovery of phytochemicals as potential Pin1 inhibitors for cancer: pharmacophore modelling, molecular docking, MM-GBSA calculations and molecular dynamics studies, Front Chem, 12 (2024) 1339891.

[49] E. Wang, H. Liu, J. Wang, G. Weng, H. Sun, Z. Wang, Y. Kang, T. Hou, Development and Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein-Ligand Binding Affinities, J Chem Inf Model, 60 (2020) 5353–5365.